

# Language Preservation 2.0

Crowdsourcing Oral Language Documentation Using Mobile Devices

---

Steven Bird

Dept of Computing and Information Systems, University of Melbourne

Linguistic Data Consortium, University of Pennsylvania

Acknowledgements:

*Florian Hanke, Haejoong Lee, Oliver Adams, Sangyeop Lee, Met Bettinson, Lauren Gawne, Katie Gelbart, Isaac McAlister*

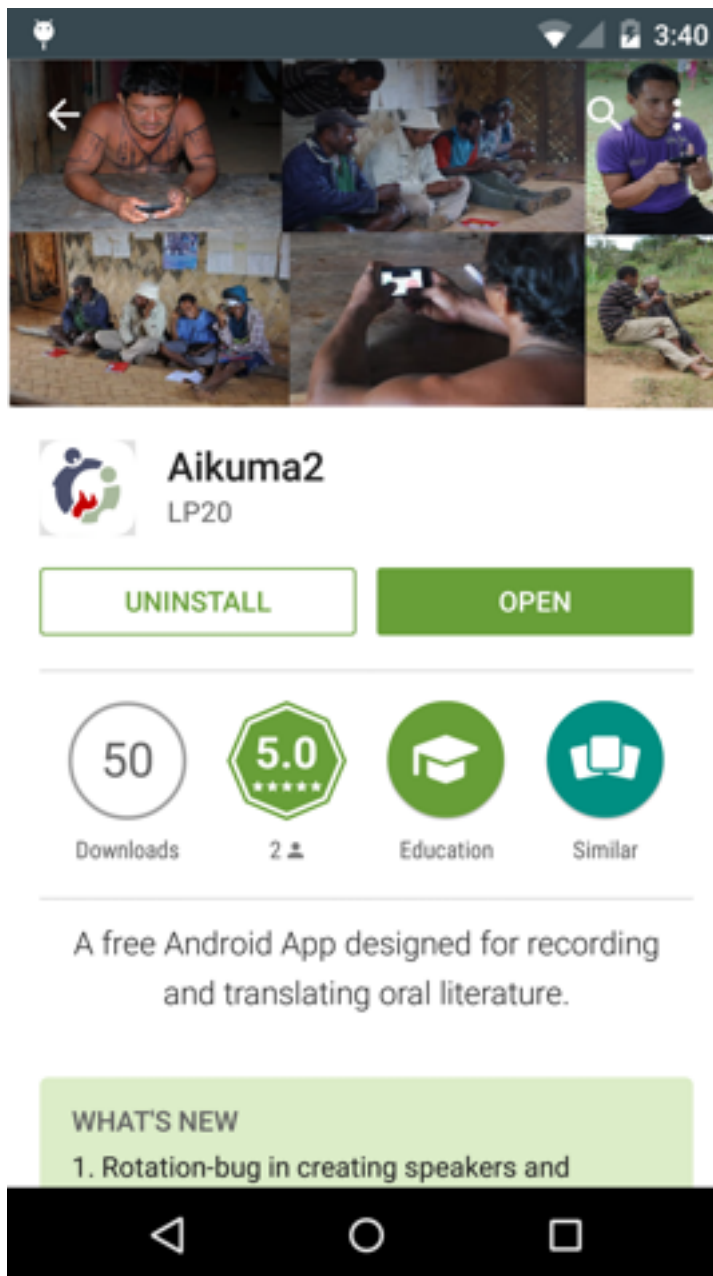
NSF DEL *Language Preservation 2.0*

ARC *Language Engineering in the Field*

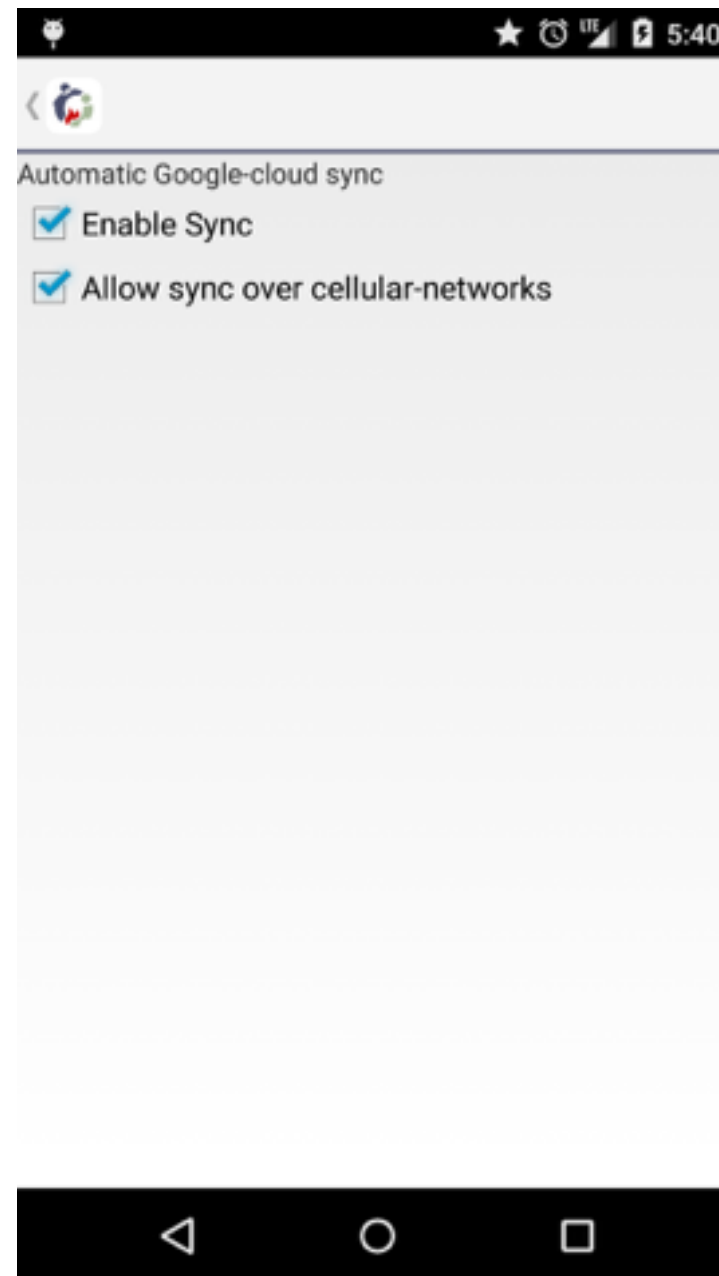


# Aikuma

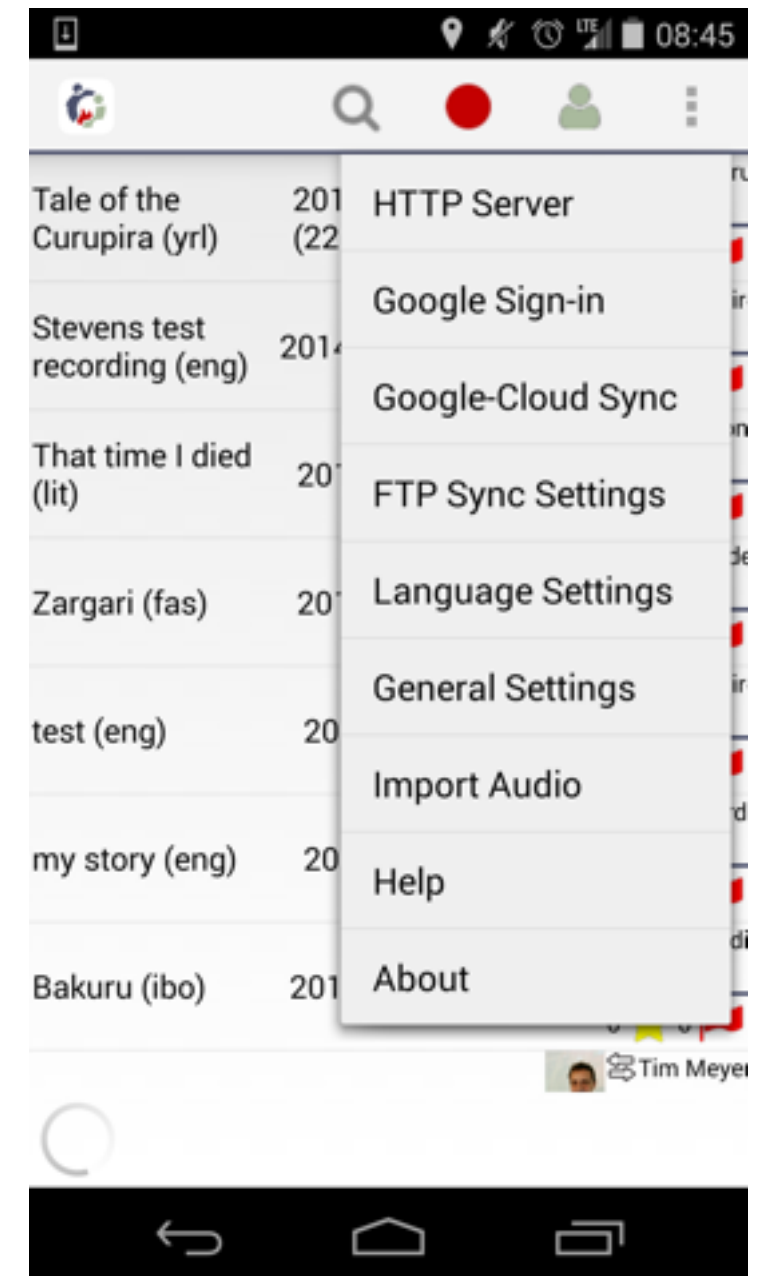
## 1. Install Aikuma2



## 2. Enable Sync



## 3. Sign-in





Aikuma = “meeting place” – Usarufa [usa]



Florian Hanke



Usarufa, PNG, 1000 spkrs > 30 years old





Aikuma: recording narratives (near field audio)

---



Temb , Brazil, 60 spkrs > 40 years old

Aikuma: oral translation (consecutive interpretation)

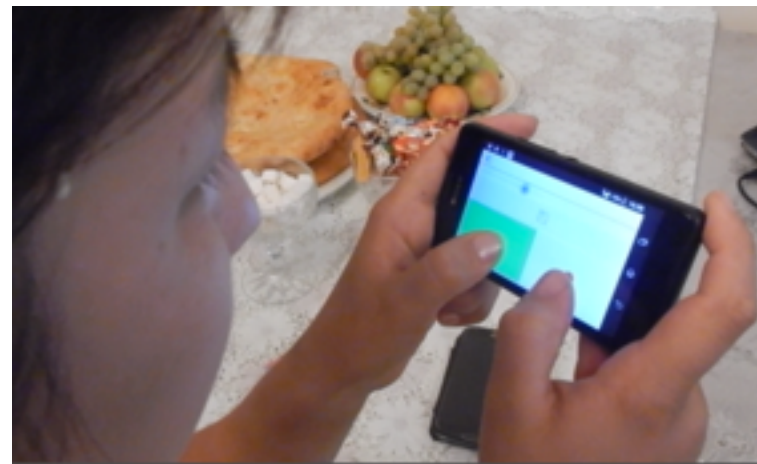
---



Tembé, Brazil, 60 speakers

# Interleaved playback

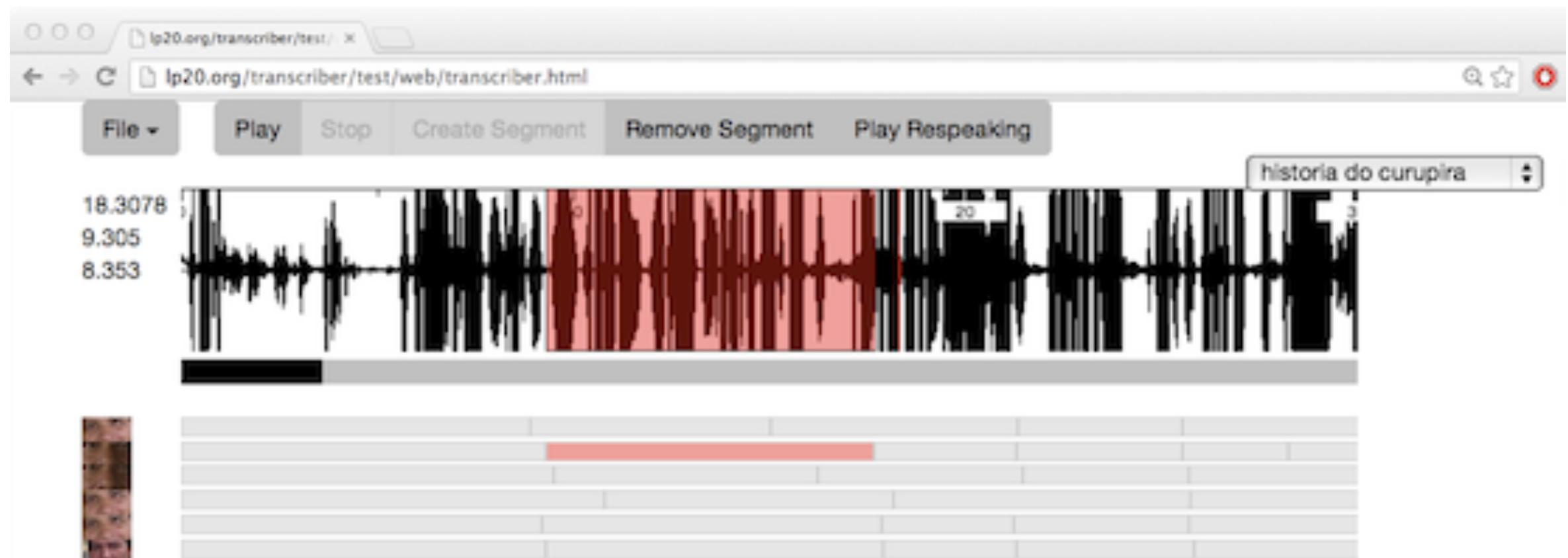
---



Karakalpak,  
Uzbekistan  
410k spkrs



# Transcription: Web app



- Certo dia um homem foi para o mato.  
One particular day a man went into the forest
- Ele gostava muito de trabalhar de tucumã, cipó e plassaba.  
He very much liked to collect tucumã, cipó and plassaba
- Ele passa durante o seu trabalho mais ou menos uma..duas ou três semanas.  
He went out to the forest more or less every one, two or three weeks
- Trabalha muito longe, quase no meio do..do centro da mata.  
He works very far away almost in the middle of the..in the center of the forest

# File Structure: Upper Level

---

aikuma/

v01/

t/

to/

**tony@gmail.com/**

items/

social/

speakers/



# File Structure: Lower Level (a)

---

tony@gmail\_com/  
items/

ualwfcixlzpu/

ualwfcixlzpu-tony@gmail.com-source.wav

ualwfcixlzpu-tony@gmail.com-source-preview.wav

ualwfcixlzpu-tony@gmail.com-source-metadata.json

ualwfcixlzpu-tony@gmail.com-comment-829120.wav

ualwfcixlzpu-tony@gmail.com-comment-829120-mapping.txt

ualwfcixlzpu-tony@gmail.com-comment-829120-metadata.json

social/

speakers/

JSLFOEPSOVKW/

JFKLEPELWXAE/

# File Structure: Lower Level (b)

---

sally@gmail\_com

items/

ualwfcixlzpu/

ualwfcixlzpu-sally@gmail.com.transcript.txt

ualwfcixlzpu-sally@gmail.com-transcript-metadata.json

ualwfcixlzpu-sally@gmail.com-respeak-493021.wav

ualwfcixlzpu-sally@gmail.com-respeak-493021-mapping.txt

ualwfcixlzpu-sally@gmail.com-respeak-493021-metadata.json

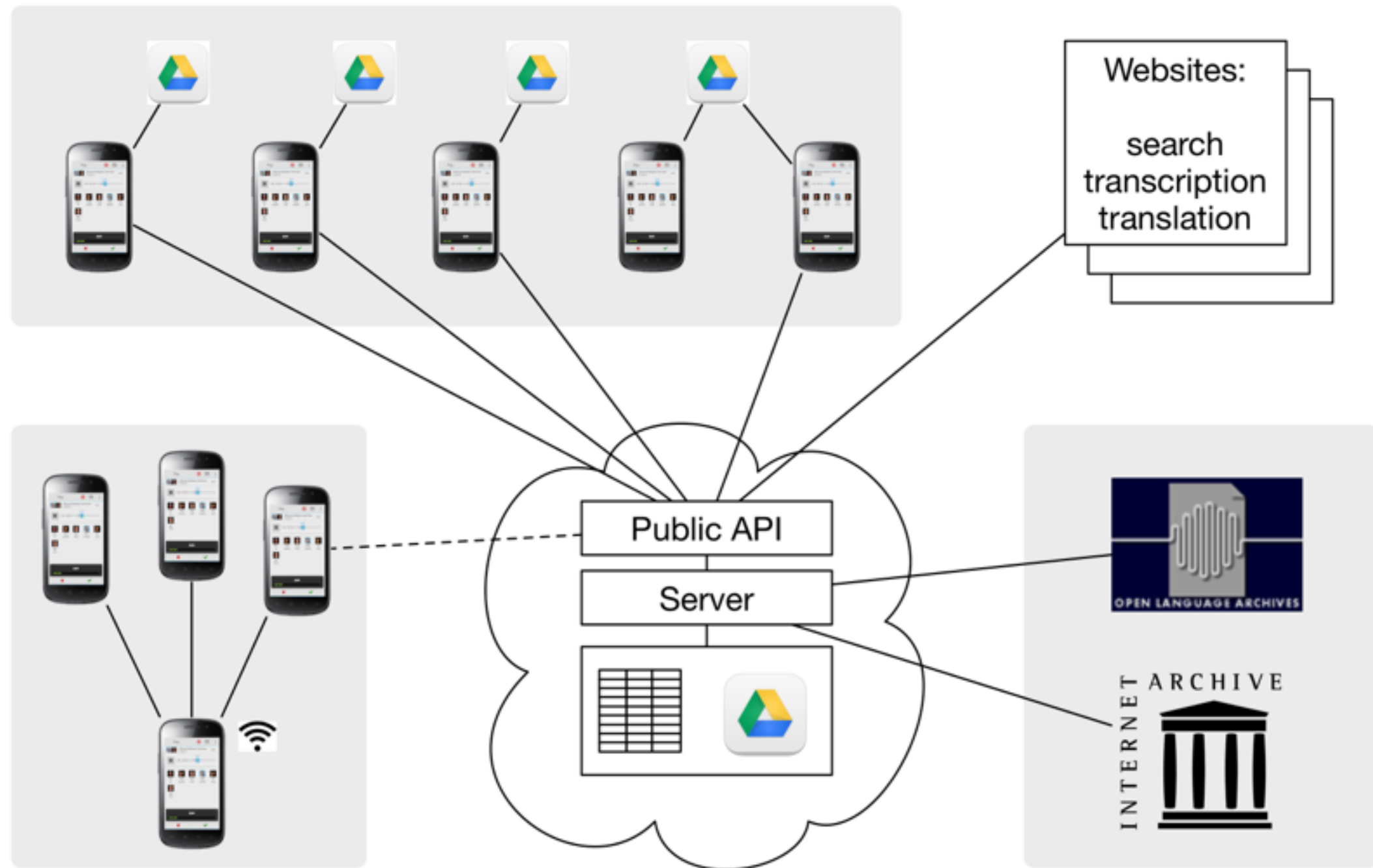
social/

ualwfcixlzpu-tony@gmail.com-source-like

ualwfcixlzpu-keren@gmail.com-interpret-139273-report

# Architecture: Networked

---



# *Thank you*

## **Get Involved**

Search for Aikuma on Google Play or on GitHub  
Email: [stevenbird1@gmail.com](mailto:stevenbird1@gmail.com)

## **References**

F Hanke & S Bird (2013). Large-scale text collection for unwritten languages. *IJCNLP*.

S Bird, L Gawne, K Gelbart, I McAlister (2014). Collecting bilingual audio in remote indigenous communities. *COLING*.

S Bird, F Hanke, O Adams, H Lee (2014). Aikuma: A mobile app for collaborative language documentation. *Workshop on the use of computational methods in the study of endangered languages*. ACL.

Fork me on GitHub